GENOME-WIDE ASSOCIATION STUDIES

Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*

Joy Bergelson and Fabrice Roux‡*

Abstract | A major challenge in evolutionary biology and plant breeding is to identify the genetic basis of complex quantitative traits, including those that contribute to adaptive variation. Here we review the development of new methods and resources to fine-map intraspecific genetic variation that underlies natural phenotypic variation in plants. In particular, the analysis of 107 quantitative traits reported in the first genome-wide association mapping study in *Arabidopsis thaliana* sets the stage for an exciting time in our understanding of plant adaptation. We also argue for the need to place phenotype–genotype association studies in an ecological context if one is to predict the evolutionary trajectories of plant species.

Adaptive walk
The evolutionary path taken by a population towards a new phenotypic optimum; it is defined by the number, phenotypic size and temporal sequence of genetic changes.

Life history
Life history traits are closely related to fitness traits, such as number and size of offspring, age at first reproduction, and reproductive lifespan and ageing.

**Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, Illinois 60637, USA. ‡Laboratoire Génétique et Evolution des Populations Végétales, FRE CNRS 3268, Université des Sciences et Technologies de Lille – Lille 1, F-59655 Villeneuve d'Ascq cedex, France. Correspondence to J.B. e-mail: jbergels@uchicago.edu*
doi:10.1038/nrg2896

Details regarding how adaptation proceeds remain elusive. From a theoretical perspective, several questions have been addressed[1–6]. How many genes are expected to be involved in a specific adaptation? Does the origin of adaptation (new mutations versus standing genetic variation) affect the adaptive walk to a new phenotypic optimum? What is the distribution of phenotypic effects that are fixed during an adaptive walk? Unfortunately, as in other long-standing debates in evolutionary ecology, arguments can flourish in the absence of data. To fill the gap between theory and data, an important goal is to identify the genetic basis of adaptive trait variation.

In plants, the identification of genes that underlie phenotypic variation can have enormous practical implications by providing a means to increase crop yield and quality in an agricultural context[7]. At the same time, the identification of ecologically important genes should help in predicting the evolutionary trajectories of plant populations[3,8–10]. *Arabidopsis thaliana* is a convenient species for these pursuits because it has a worldwide distribution and, as such, encounters diverse ecological conditions[9,11–14], leading to adaptive variation for many morphology, life history and other fitness-related traits[15]. During the past two decades, molecular tools have been developed to assist in the mapping of quantitative trait loci (QTLs) in experimental populations, but these tools remain laborious[16]. Recently, the first study of genome-wide association (GWA) mapping in plants was reported[17], bringing a breath of fresh air to the area of gene discovery. The high resolution conferred by GWA mapping facilitated mapping of the genetic bases of 107 diverse phenotypes, including flowering time, pathogen resistance, seed dormancy, ionomics and vegetative growth. Long considered the privilege of human mapping studies, GWA mapping has now emerged as a powerful alternative approach to finely dissect the intraspecific genetic variation that underlies phenotypic variation in plants[18–20].

Here we describe the connections among long-established strategies (such as traditional linkage mapping), recently developed approaches (such as GWA mapping) and upcoming methods (such as nested association mapping (NAM)) for finely mapping QTLs underlying natural variation. We review several powerful GWA mapping approaches and analytical methods that have been developed, as well as the available genotypic and phenotypic resources that are linked to the approaches. Because genetic variation is exposed to natural selection in contrasting ecological habitats, we emphasize the importance of ecological context. First, the spatial and temporal scale at which selection acts will determine the appropriate populations for GWA mapping[21]. Second, the cues perceived by a plant are far more complex, and not well captured, by simple growth-chamber conditions. This highlights the need to measure phenotypes in realistic conditions[22,23]. Third, the heterogeneity of the habitats encountered by *A. thaliana* suggests that experiments designed to phenotype plants in multiple locations will provide more robust results than

Table 1 | **Advantages and drawbacks of methods for identifying the genetic basis of complex traits in *Arabidopsis thaliana***

Methods	Starting year	Advantages	Drawbacks	Refs
Traditional linkage mapping, that is, QTL mapping	1992	• No population structure effect • Identification of rare alleles • Few genetic markers required for a complete genome scan	• Coarse mapping • Limited genetic diversity • Not possible to distinguish between pleiotropic and physically close genes	30
Association mapping with candidate genes	2002	• Fine mapping	• Requires detailed knowledge of the biochemistry and genetics of the trait under study • Approach is biased for previously identified genes	42,146, 147
GWA mapping at the species scale	2005	• Fine mapping (blind approach) • Detection of common alleles	• False positives due to population structure • False negatives after controlling for population structure • Reduced power to detect rare alleles or weak-effect alleles • Genetic and allelic heterogeneity	17,46
Dual linkage–association mapping at the species scale (FIG. 2)	2007	• Fine mapping (blind approach) • Identification of false positives and false negatives	• Phenotyping of several thousands of individuals • Numerous traditional linkage mapping populations required • Genetic and allelic heterogeneity	23,49, 52
GWA mapping in regional mapping populations	2010	• Fine mapping (blind approach) • Diminished population structure effect • Detection of genes involved in local adaptation	• Potential for limited phenotypic variation • Increased linkage disequilibrium: less precise than using a worldwide sample	21,44, 114
NAM at the species scale	Ongoing	• Fine mapping (blind approach) • Identification of false positives and false negatives • High-density genotyping of a small number of founders lines (<30)	• Importance of the crossing schemes and the number of founders • Phenotyping of several thousands of individuals • Genetic and allelic heterogeneity	63,64, 148

GWA, genome-wide association; NAM, nested association mapping; QTL, quantitative trait locus.

Quantitative trait locus
Genomic region containing one or more genes that affect the variation of a quantitative trait.

Genome-wide association
Whole-genome scans that test the association between the genotypes at each locus and a given phenotype.

Seed dormancy
Mechanism that prevents seed germination, even under conditions that promote germination.

Ionomics
The study of the composition of mineral nutrients and trace elements in living organisms.

Genotype–environment interaction
An effect of a locus that changes in magnitude or direction across environments.

Trade-off
Negative genetic and phenotypic correlation between two traits arising from the need of the individual to allocate resources to alternative functions.

will those designed to phenotype plants in only one location, while also offering insights into the genetic bases of genotype–environment interactions (G×E interactions)[24,25]. Last, in nature there are a multitude of selective pressures that simultaneously act on individuals. This should lead to selection for a global phenotypic optimum that results from trade-offs among specific traits[26]. We argue that the adaptive value of a specific trait is best understood in the context of other phenotypic traits, when its relative contribution to fitness is known.

Next-generation sequencing (NGS) technologies[27–29] will additionally facilitate access to the causal polymorphisms that underlie natural variation of complex traits. This is clearly an exciting time to map the genetic bases of complex traits in *A. thaliana* and put them in the context of ecology and adaptation in nature. In this Review, we first assess alternative methods for identifying natural alleles that control quantitative traits, addressing them chronologically according to their use in *A. thaliana* (TABLE 1). We then outline the prospects for introducing ecological approaches to the genetic analyses.

Traditional linkage mapping

Based on a genetic map, traditional linkage mapping (also known as QTL mapping) in *A. thaliana* refers to the use of experimental populations (FIG. 1; TABLE 2) ranging from classical F2 populations[30] to the more recently developed multiparent advanced generation intercross (MAGIC) lines[31]. Recombinant inbred lines (RILs) remain the most popular experimental populations in *A. thaliana*: as these populations are almost completely homozygous, they allow one to replicate genotypes within an experiment and/or among several environmental conditions. More than 60 such RIL families have already been developed. RILs typically define QTL regions of a few megabases covering thousands of genes[32,33], although the resolution can be as high as 300 kb (~50 genes) for the MAGIC lines[31]. A major drawback of this approach, therefore, is that the resultant mapping is coarse (TABLES 1,2). Three options can be envisaged to resolve this issue. First, after a QTL has been localized to a relatively narrow region (3 cM or less, for example (see REF. 34)), fine mapping and cloning of the QTL can be carried out, typically using near isogenic lines (NILs) or heterogeneous inbred families (HIFs)[35,36]. Second, QTL mapping can be complemented with microarrays or sequence prediction for inactivated genes within QTL intervals[37,38]. Third, a promising alternative to fine mapping in QTL regions involves direct sequencing of segregating populations to identify causative mutations, as first demonstrated for induced mutants in *A. thaliana*[39,40] and then implemented for quantitative traits in yeast[41].

Unlike GWA mapping, traditional linkage mapping is useful for identifying rare alleles and is not subject to the effect of population structure (see the later subsection 'GWA studies: the disadvantages'; TABLES 1,2). The genetic bases that are identified by QTL mapping, however, are specific to the parental lines of the experimental segregating populations and may not be representative of the genetic variation on which natural selection acts.

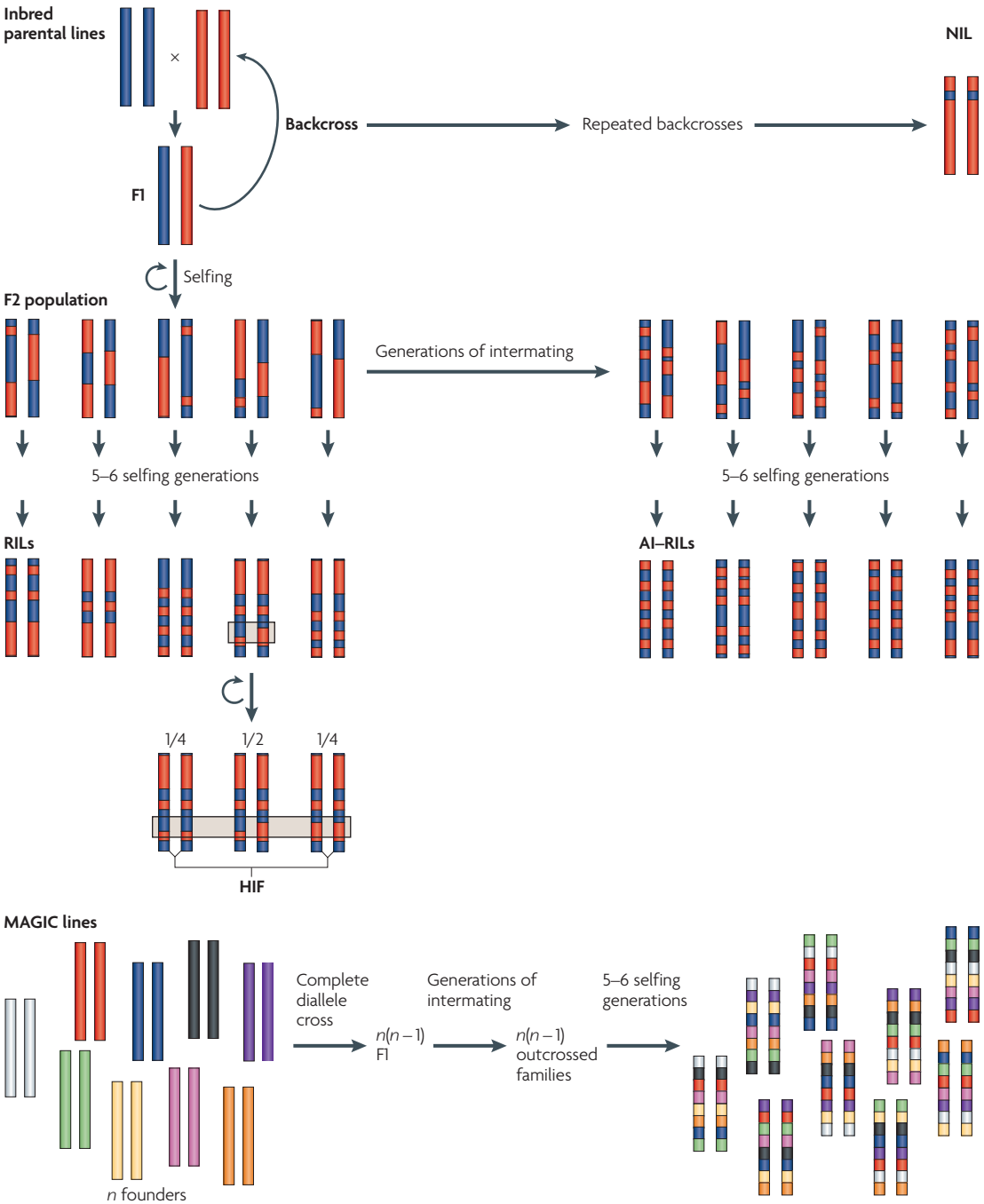


Figure 1 | **Linkage mapping populations in *Arabidopsis thaliana*.** The mapping resolution and the genetic diversity in the linkage mapping populations will depend on the number of founders, generations of intermating and generations of selfing. See TABLE 2 for the advantages and drawbacks of each mapping population. AI-RILs, advanced intercross–recombinant inbred lines; HIF, heterogeneous inbred family; MAGIC lines, multiparent advanced generation intercross lines; NIL, near-isogenic line; RILs, recombinant inbred lines.

Genetic map
Representation of the position of genetic markers relative to each other, with distances between loci expressed in terms of recombination frequency.

Recombinant inbred lines
Quasi-homozygous lines produced from an initial cross between two individuals, followed by six to eight generations of selfing.

Population structure
Differentiation in allele frequencies among multiple populations.

Linkage disequilibrium
Nonrandom allelic association such that two alleles at two or more loci are more or less frequently associated than predicted by their individual frequencies.

Association mapping, GWA studies and NAM

Association mapping appeared next as an alternative for fine-mapping genomic regions associated with phenotypic variation; this method has taken both candidate gene and genome-wide approaches. The candidate gene approach is especially useful in non-model plant species; however, it requires detailed knowledge of the biochemistry and genetics of a trait[7], making it difficult to apply even in a well-studied species such as *A. thaliana*[42]. By contrast, the genome-wide strategy allows one to search blindly for genomic regions that are associated with a trait of interest[43]. GWA mapping uses natural linkage disequilibrium (LD) to identify polymorphisms that are associated with phenotypic variation. As GWA studies take advantage of recombination events that have accumulated over thousands of generations[9,18], the resolution

Table 2 | **Advantages and drawbacks of linkage mapping populations in *Arabidopsis thaliana***

Mapping material	Advantages	Drawbacks	Time (generations)	Refs
Backcross	• Detecting genetic basis of heterosis*	• Low mapping resolution • Limited genetic diversity	2	149,150
F2 population	• Estimation of QTL dominance	• Genotyping individuals for each phenotyping experiment • Limited genetic diversity	2	52,151
RILs	• Genotyped once • Unlimited replicates	• Limited genetic diversity	7–8	33,58
AI-RILs	• High-resolution mapping • Genotyped once • Unlimited replicates	• Limited genetic diversity	10	152,153
MAGIC lines	• High-resolution mapping (up to 300 kb) • Increased genetic diversity • Genotyped once • Unlimited replicates	• Genetic and allelic heterogeneity	10	31
NILs	• Single introgression segment in homogeneous genetic background • Increased power to detect small-effect QTL • Unlimited replicates	• Time consuming: size of the introgression segment will depend on the number of backcross generations • Limited genetic diversity	>6	36
HIFs	• Single introgression segment in heterogeneous genetic background • Increased power to detect small-effect QTLs • Increased power to detect epistasis • Unlimited replicates • The same genomic region covered by independent HIFs	• Limited genetic diversity	9–10	35,154

*Heterosis is the equivalent to hybrid vigour; superiority in one or more phenotypes of the hybrid individual over the parents. AI-RILs, advanced intercross–recombinant inbred lines; HIF, heterogeneous inbred family; MAGIC lines, multiparent advanced generation intercross lines; NIL, near-isogenic line; QTL, quantitative trait locus; RILs, recombinant inbred lines.

to fine map can be greatly enhanced relative to RILs. *A. thaliana,* in particular, has LD that extends for roughly 10 kb[44], which is a nearly ideal distance for mapping — that is, it extends up to the gene level, so there is no need to develop extensive SNP-tiling arrays.

GWA studies: the advantages. A meta-analysis of GWA studies for 107 phenotypic traits in *A. thaliana* was recently carried out[17], and all of the genetic and phenotypic resources are publically available. In the study, 76 to 194 accessions, that is, genetic lines sampled in natural populations, were phenotyped for traits related to flowering time, developmental characteristics, biotic resistance and/or ionomics. These accessions, which are propagated as homozygous lines, have been genotyped using AtSNPtile1 (REFS 44,45), a custom Affymetrix SNP chip containing almost 250,000 known non-singleton SNPs. The ability of GWA mapping in *A. thaliana* to identify the genetic basis of various phenotypic traits has been demonstrated in three main ways. First, GWA mapping successfully identified resistance genes that were already validated as those underlying resistance to pathogens in *A. thaliana*[46]. For example, the *RESISTANT TO PSEUDOMONAS SYRINGAE 5* (*RPS5*) *R* gene — which is known to recognize the *avrPphB* avirulence gene in the bacterial pathogen strain *Pseudomonas syringae*: Pst DC3000 (REF. 47) — was detected as a single peak of association. The same is true for other known *R* genes, such as *RESISTANCE TO P. SYRINGAE PV MACULICOLA 1* (*RPM1*)[48]. Similarly, association peaks related to qualitative resistance to the downy mildew agent *Hyaloperenospora arabidopsidis* ex *parasitica* (*Hpa*) overlapped with known RPP (resistance to *Hpa*) loci[49]. Second, the main association peak identified by GWA mapping for leaf necrosis in a set of 96 accessions was located within the ACCELERATED CELL DEATH 6 (ACD6) locus, which was functionally validated as the main determinant of natural variation for premature leaf death[50]. Third, based on previous knowledge of the very detailed genetic network of flowering time, enrichment of *a priori* candidate genes has been found for several traits related to floral transition[17,23]. Although this enrichment of *a priori* candidate genes is encouraging, only functional validation will prove that the genes related to flowering transition that were identified under association peaks are true positives.

GWA mapping in humans generally requires thousands of genotyped individuals to account for a small fraction of the genetic variation of complex traits[51]. Even with fewer than 200 genotyped accessions, strong associations have been found in *A. thaliana*, suggesting the occurrence of common alleles of major effect at the species scale[17]. As GWA mapping is a blind approach, it also facilitates the identification of new regions containing no *a priori* candidate genes[17,23], potentially enhancing our knowledge of genetic networks related to complex traits.

GWA studies: the disadvantages. GWA mapping in *A. thaliana* suffers from two major limitations. First is the problem of false positives due to population structure (FIG. 2). Population structure may be a problem that is especially great when both phenotypic and

SNP-tiling array
A microarray platform combining SNP genotyping and whole-genome tiling; it contains probes for each allele and each strand of several thousands of SNPs.

Non-singleton SNP
A SNP polymorphism that is present in at least two individuals.

genetic differentiation vary with geographic distance[52]. Statistical methods to control for population structure can reduce the inflation of false-positive associations (see the 'Statistical analyses for GWA mapping' subsection) but may also introduce false negatives (rarely considered in GWA studies in humans); that is, causative genetic markers may be lost when applying GWA methods that control for population structure (FIG. 2). One potential solution is to carry out GWA mapping on a less structured sample of accessions; however, this alternative is not feasible when the phenotypic variation occurs on the scale of the species[21]. In such cases, a combination of traditional linkage mapping and GWA mapping may be a better alternative for reducing the rate of false positives[49,53] and for detecting false negatives[52] (FIG. 2). Dual linkage and association mapping was recently shown to outperform each method in isolation when applied to flowering time data for *A. thaliana* grown in the field[23]. For the 50 best-associated SNPs, the enrichment ratio in *a priori* candidate genes almost doubled when considering candidate genes overlapped by QTLs detected using RILs relative to candidate genes only (7.4 versus 4.1, respectively). This dual mapping strategy estimated that GWA analysis alone led to a false-positive rate of 40% and a false-negative rate of 24%.

Second, genetic heterogeneity and/or allelic heterogeneity may interfere with the detection of SNPs linked to phenotypic variation (FIG. 3). It is well known that different combinations of genes can lead to the same phenotype[54]. For example, the genetic bases of the coat colour that confers a selective advantage of crypsis[55] in the coastal beach mouse *Peromyscus polionotus* differs between populations in the Mexican gulf and on the east coast of Florida[56]. In several plant species, different QTLs[23,57,58] and/or different alleles at the same QTL[59–62] are responsible for an early-flowering trait. As a first step to control the effects of genetic and allelic heterogeneity in *A. thaliana*, >1,100 *A. thaliana* lines have been collected and genotyped using the Affymetrix 250K SNP-tiling array, AtSNPtile1. This set, called the RegMap lines, covers much of the geographical range of the species but with particularly strong representation of accessions from Sweden, the United Kingdom and France (J.B., J. Borevitz and M. Nordborg, unpublished data). The comparison of GWA mapping results among subsets of this collection will reveal the extent of genetic and allelic heterogeneity in *A. thaliana*. Although they capture much of the genetic variation in the species, the RegMap lines are nonetheless geographically limited, and more extensive sampling in additional geographic regions would be desirable.

Nested association mapping. NAM, which was originally developed in *Zea mays*[62,63], is a promising method that is currently under development for fine-mapping QTLs in *A. thaliana*. NAM takes advantage of both historic and recent recombination events to combine the advantages of traditional linkage mapping (that is, low marker-density requirements and high allele richness) and association mapping (that is, high mapping resolution and high statistical power), while being less susceptible to false positives and false negatives[64]. The 250K SNP genotyping of many *A. thaliana* accessions that serve as parents in RIL populations or MAGIC lines will soon enable the NAM strategy to be undertaken. In practice, this will involve projecting the genetic information from parental lines onto the experimental populations used in a traditional linkage mapping study. The joint analysis of data sets from natural accessions and NAM populations should greatly increase our power to finely map genomic regions associated with phenotypic variation, although statistical analyses adapted to the hierarchical design of NAM remain to be developed.

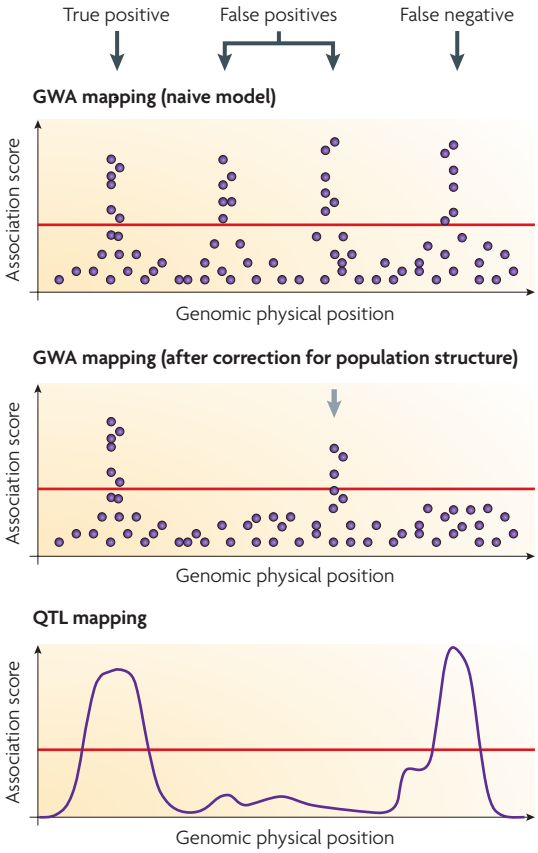


Figure 2 | **Advantages of combining association and traditional linkage mapping methods.** Dual linkage–association mapping allows true positives and false negatives to be distinguished from false positives. True positives are causative SNPs that have been detected by genome-wide association (GWA) mapping and are overlapped by quantitative trait locus (QTL) regions. Population structure corrections highlight false positives that correspond to false phenotype–genotype associations. Because statistical methods that control for population structure only reduce (but do not abolish) the inflation of false positives, false positives may remain (grey arrow). In such cases, the remaining false positives are not validated by QTL regions, demonstrating the added value of QTL mapping in the detection of true positives. False negatives are causative SNPs that are lost as an artefact of population structure corrections but can be validated by QTL regions. The horizontal red line indicates the significance threshold for a phenotype–genotype association.

Genetic heterogeneity
The same phenotypic value caused by different mutations at different genes.

Allelic heterogeneity
The same phenotypic value caused by different mutations at the same gene.

Crypsis
Capacity of an organism to avoid detection by other organisms by blending into the environment.

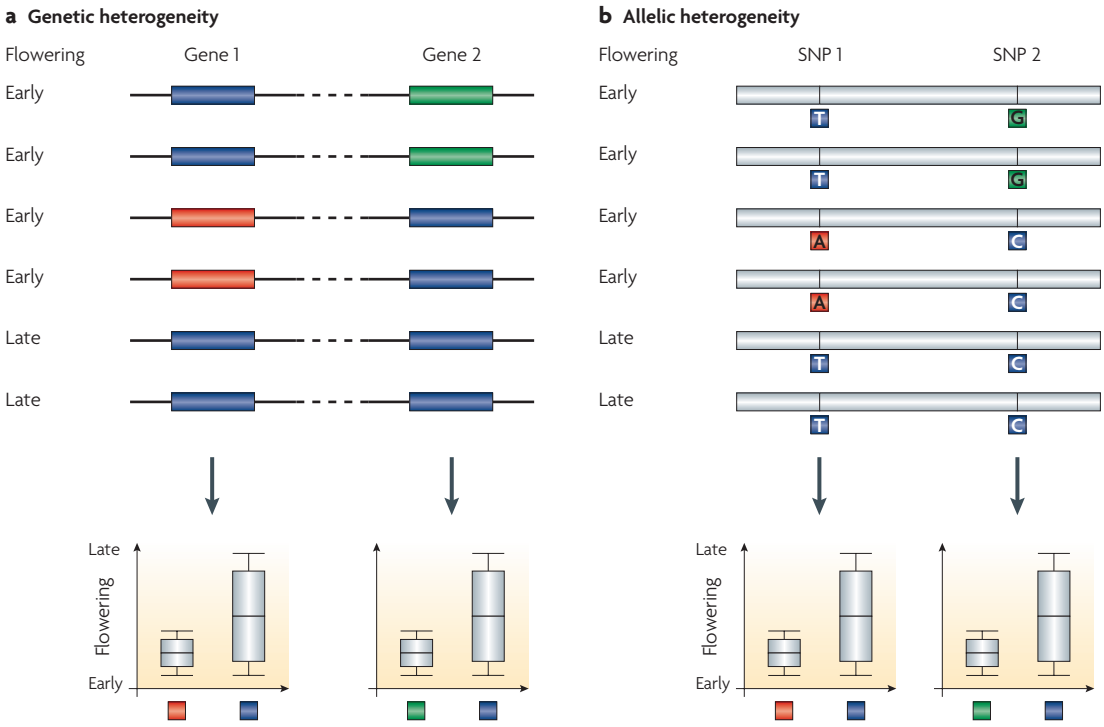


Figure 3 | **Genetic and allelic heterogeneity. a** | Genetic heterogeneity. When alternative genes lead to the same phenotype, genetic heterogeneity can impede detection of the genes that underlie natural phenotypic variation. Here, early flowering occurs through different quantitative trait loci (QTLs), that is, genes 1 (red allele) and 2 (green allele). **b** | Allelic heterogeneity. Two alleles at the same gene (T→A and C→G mutations) may confer a similar phenotype, such as early flowering. Box plots associated with genetic and allelic heterogeneity are represented in the lower panel for each polymorphic gene and polymorphic allele, respectively.

GWA tools: marker types. SNP markers are increasingly popular for mapping because of their high frequency in the genome[29]. SNP-tiling arrays (AtSNPtile1) containing probe sets for 248,584 SNPs have been designed using the complete genome sequences of 20 natural accessions[44] that represent the maximal genetic diversity among a set of 95 worldwide accessions[65]. Given the relatively small genome size of *A. thaliana*, this Affymetrix genotyping array provides, on average, one SNP every 500 bp[44], which is more than adequate coverage for GWA mapping. Indeed, even though LD extends an average of 10 kb among worldwide accessions of *A. thaliana*, the coverage afforded by this array is sufficient to accommodate high variability in LD across the genome[44]. For example, LD is less extensive around loci that have experienced balancing selection for long evolutionary times, such as those encoding pathogen resistances[66–68].

SNP markers represent only a fraction of the available genetic polymorphisms. The ongoing 1001 Genomes sequencing project will permit access to most DNA polymorphisms in *A. thaliana*[69], including structural variants such as copy number variation (CNV) or insertions-deletions (indels). Indels contribute to phenotypic variation in *A. thaliana* for traits such as flowering time[59,60,70] and resistance to herbivory and pathogens[67,71]. Initial analysis of ~200 German and Swedish whole-genome sequences revealed 5 million SNPs (M. Nordborg, personal communication). Information on the full genome should facilitate direct access to causal variations, greatly facilitating a mechanistic understanding of functional polymorphisms. Although it might seem that consideration of whole-genome sequences would vastly increase the computational time required for testing associations, many SNPs will be linked, and this should enable statistical analyses to be hierarchical based on a limited set of SNPs.

Epigenetics is also well known to shape phenotypic variation and should be considered in efforts to understand the evolution of complex traits[72,73]. In a genome-wide survey of loci on chromosome 4, DNA methylation was found to be highly polymorphic among a set of 96 natural accessions of *A. thaliana*[74]. Epigenetic variation can account for up to 30% of the variation in flowering time and plant height in *A. thaliana*[75]. Given the recent technical revolution, epigenome characterization at single-base-pair resolution can be envisioned[76–78]. Indeed, genome-wide scans of DNA methylation are underway in natural accessions of *A. thaliana*. The combination of genotypic and epigenetic information will help to tease apart the effect of DNA sequence variants from that of DNA methylation variants. The development of epigenetic RILs (epiRILs)[75] will also enable the combination of both linkage and association mapping at the DNA methylation level.

Balancing selection
Evolutionary processes that maintain genetic diversity within a population for longer than expected under neutrality. Processes include heterozygote advantage, frequency-dependent selection and variation of fitness in space and time.

Epigenetic RILs
Quasi-homozygous lines that are almost identical at the genetic level but segregate at the DNA methylation level. EpiRILs are produced from an initial cross between two individuals with few DNA sequence differences but contrasting DNA methylation profiles, followed by six to eight generations of selfing.

Given the ever-increasing genetic and epigenetic information for each *A. thaliana* accession, GWA studies will soon suffer from the problem of large dimensionality of polymorphisms. Because the inclusion of additional information may generate more false-positive associations between phenotype and polymorphic markers, statistical tools to appropriately reduce the false-positive rate are in demand.

Statistical analyses for GWA mapping. The effect of SNPs on phenotypes can be tested by using one of several models: non-parametric methods, such as the Wilcoxon rank-sum test for ordered categorical and quantitative phenotypes, or Fisher's Exact Test for binary phenotypes. However, these are relatively naive models because they fail to take into account the confounding effects of population structure. Numerous methods have been developed to account for confounding due to population structure (reviewed in REF. 79). The EMMA[80] software includes a matrix of genotype similarity among the accessions in its mixed linear model (MLM); this matrix has been shown to efficiently correct for the effects of population structure in *A. thaliana*[17], suggesting that the structure of the kinship matrix may well represent both population structure and cryptic relatedness[81]. One limitation of the MLM might be the dependence of associations on minor allele frequencies: strong phenotypic associations are more readily detected when the minor allele frequency is low[17]. Although rare alleles may, at times, be associated with strong phenotypic effects[82], much of this enrichment is likely to be spurious[17].

Given the increasing number of individuals genotyped (with increasing numbers of markers), various methods have been developed to reduce computing time while maintaining or improving statistical power to control for population structure and cryptic relatedness. Such methods include 'compressed MLM' and 'population parameters previously determined (P3D)'[83] — both of which are implemented in the software TASSEL[84] — and a similar variance component approach that is implemented in the EMMAX software[85]. These methods involve first estimating the contribution of population structure to the phenotype using a variance decomposition model. The resulting genetic variance and residual variance values are then kept fixed in a model that tests an association between each marker and the phenotype.

Many phenotypic traits might be structured as a network, time series or hierarchy (BOX 1). Numerous GWA mapping extensions are underway to take advantage of such phenotypic structure for increasing the detection of associated genome variations[86]. Because the causal allele for the same phenotype might be different among populations[12,56,87], multi-task regularized regression has also been used to find causal loci in multi-population GWA mapping[88] and to reduce the rate of false positives due to population structure. Such structured association-mapping algorithms are often publically available on platforms such as GenAMap.

Functional validation

Although GWA mapping has greatly enhanced our ability to fine-map the genomic regions that are linked to natural variation, functional validation remains the gold standard for identifying causative polymorphisms. This is facilitated by the impressive genetic resources in *A. thaliana*[16,34,89], such as non-targeted random disruption or alteration (ethyl methanesulfonate (EMS)- and transposon-mediated mutagenesis, T-DNA mutants and unimutant collection) and specific targeted disruption or alteration (gene silencing by amiRNA), which allow quantitative complementation and quantitative knockdown, respectively.

Nevertheless, we must keep in mind that QTLs that are detected by either QTL mapping or GWA mapping will often result from allelic variation that cannot be captured by knockout or knockdown lines. In addition, QTLs that are detected in field experiments typically explain less than 10% of phenotypic variation[23,90]. For both of these reasons, it is important to create isogenic material against which the effects of particular genes can be compared while also controlling for effects of genetic background and chromosomal location. As an example, the use of a Cre–*lox* system facilitated the detection of small (9%) differences in seed production that were associated with the presence or absence of the *RPM1* pathogen-resistance gene under field conditions[91]. Extensions of this technology to allow consideration of allelic series will prove to be useful but have not yet been applied to dissecting QTLs[92].

To date, more than 30 genes involved in natural variation of complex traits have been functionally validated in *A. thaliana*[93]. However, functional validation is still lacking for a range of quantitative traits that are thought to be strongly related to plant fitness, such as the duration of the reproductive period[94] or disease resistance and tolerance to pathogens[95,96]. Also noteworthy is the absence of functional validation of ecologically important genes as scored under field conditions. Because natural selection acts in nature, where the environment and associated cues are complex, we argue that both GWA mapping and functional validation under natural conditions will be crucial for understanding adaptive evolution.

Adding ecology to association mapping

Geographical scale of adaptation. The maintenance of phenotypic diversity and life history evolution will largely depend on the scale of environmental heterogeneity[97,98]. As a consequence, the scale at which GWA mapping should be performed will depend on the scale at which natural variation is observed, which in turn depends on the ecological factors acting as selective pressures. Bolting time, for example, is correlated with latitude in *A. thaliana,* making climatic variables plausible ecological factors acting across the range of the species[99,100]. Other selective pressures, such as attack by natural enemies, soil composition and interspecific competition, may well be heterogeneous among geographically close populations, and even among individuals within a population[101–103]. In addition to geographical variation is the role of temporal variation, which affects the recruitment of adaptive alleles[104,105]. The environmental grain might thus differ among phenotypic traits across spatial and temporal scales[106–108] and clearly needs to be considered in the design of regional or local mapping populations.

Non-parametric methods
Statistical methods, also called distribution free methods, that are not based on a normal distribution of data.

Mixed linear model
Statistical model containing both fixed effects and random effects.

Multi-task regularized regression
Joint association analysis of multiple populations with a multi-population group lasso using L_1/L_2 regression.

T-DNA
Transferred DNA of the tumour-inducing (Ti) plasmid of some bacterial species into the nuclear DNA genome of the host plant.

Unimutant collection
A collection of 31,033 publically available homozygous T-DNA insertion lines in *Arabidopsis thaliana* representing 18,506 individual genes; produced by the Salk Institute.

AmiRNA
Artificial microRNAs that target specific genes for silencing.

Cre–*lox*
Transgenic technology creating isolines with identical genomes, except for the gene of interest. The resulting paired isolines are created by first introducing the gene of interest with a selectable marker into the genome and then excising the gene of interest. Modifications of this approach can be used to create allelic series.

Environmental grain
The scale of temporal and spatial environmental variation that is perceived by an organism.

Projects that aim to describe the environmental grain of diverse selective pressures would be useful, especially if they emphasize factors such as biotic interactions[22], which are poorly studied in natural populations of *A. thaliana* but are well known to influence the evolutionary trajectories of populations in other plant species[109,110]. Soil composition is another important factor that may drive adaptive responses in plants[111,112]. Indeed, a short life cycle emerged as an adaptive response to high concentrations of phosphate, as experimentally validated in *A. thaliana*[113]. Furthermore, natural populations of *A. thaliana* associated with coastal and saline soils in Europe were recently found to be enriched for a weak natural allele of *HIGH-AFFINITY K+ TRANSPORTER 1;1* (*HKT1;1*), which confers elevated salinity tolerance[114].

Box 1 | **Structured phenotypic traits**

Many phenotypic traits can structured as a hierarchy, time series or network.

Hierarchy

Phenotypic variation of trait 1 may be decomposed by variation that is observed in traits 11 and 12, themselves decomposed by variation that is observed in intermediate phenotypes (see the figure, part **a**). For example, the life cycle in annual plants may be decomposed into a vegetative phase and a reproductive phase. The vegetative phase is composed of the time interval between sowing and bolting and the interval between bolting and flowering. The reproductive phase is composed of the flowering period and the seed maturation period. Note that the flowering and maturation periods may overlap.

Time series

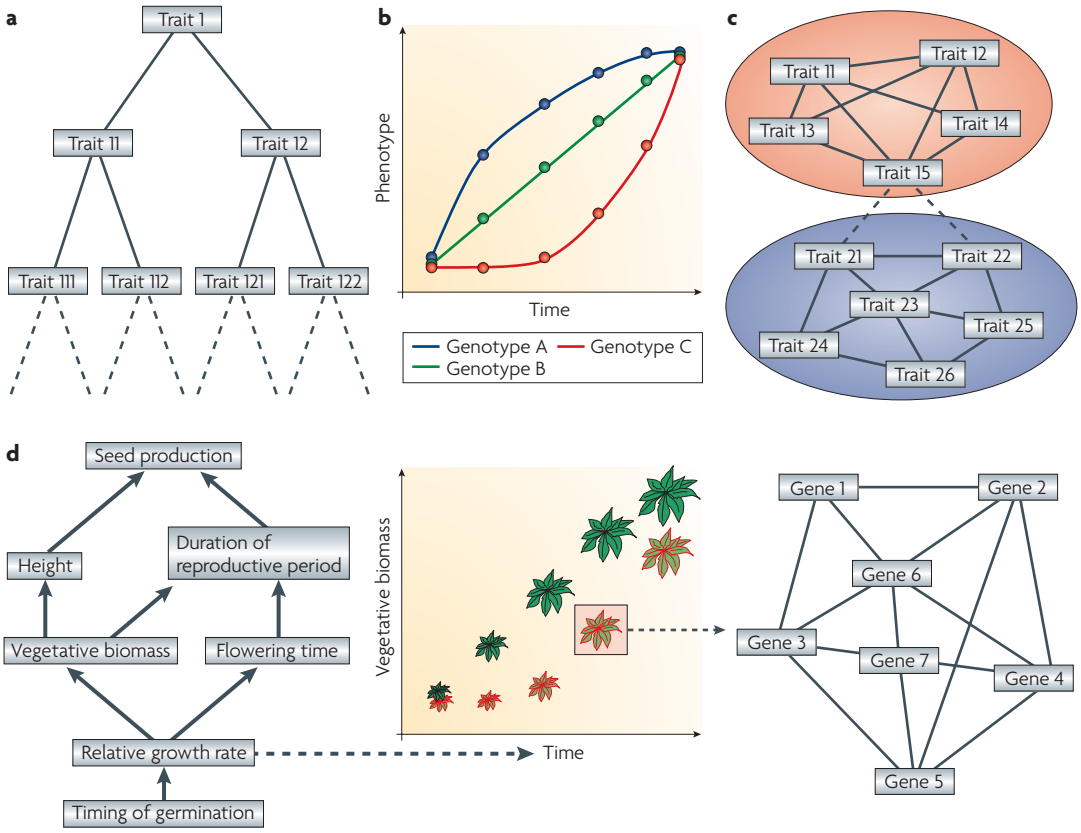
In this case genotypes are measured for the same phenotype at equally spaced, discrete time intervals (see the figure, part **b**). Time series analysis deals with the non-independence of data points taken over time. Examples of time series analysis in *Arabidopsis thaliana* include studies of disease symptoms, aerial biomass growth, root growth and cold tolerance.

Quantitative trait network

Gene expression, primary and secondary metabolite profiles, composition of mineral nutrients and trace elements could be studied as quantitative trait networks. Part **c** shows the connection between two sub-networks, each corresponding to a group of intercorrelated traits.

Interlink among hierarchy, time series and network

Seed production results from a combination of morphological, phenological and life history traits (see the figure, part **d**, left). A component of the hierarchy, relative growth rate, is estimated by scoring the vegetative biomass at successive times (middle). Vegetative biomass estimated at a specific time results from the intercorrelated expression of many genes (right).



A benefit of studying adaptation in plants is that they stand still, and because the collection site of many accessions, including all RegMap lines, is known, it is easy to envision scans for genes associated with particular environmental variables. Several such studies are currently underway and are likely to produce a rich list of candidate genes for ecological testing.

Complex environmental cues. Consistent with observations in other plant species[115], QTL mapping analyses in *A. thaliana* have revealed different QTLs for the same traits measured in greenhouse conditions and in common gardens[23,90,116,117]. The high resolution conferred by GWA mapping in *A. thaliana* strengthens this observation. Only two out of 25 candidate genes associated with flowering time when measured under field conditions have also been proposed as candidate genes for flowering-time phenotypes in GWA mapping studies scored under greenhouse conditions[23]. In a natural setting, plants are exposed to a greater range of day lengths and greater daily fluctuations in temperature, humidity and light quality than are typically encountered in the greenhouse. As a consequence, many circadian clock-related genes entrained by photoperiod and thermocycles have been detected by GWA mapping for flowering time scored in ecologically realistic conditions[23], but not in the greenhouse.

Recent QTL mapping studies of flowering time in *A. thaliana* have attempted to simulate outdoor climatic conditions in growth chambers by varying photoperiod and temperature over time[118–120]. Although this is a good first step, these studies used climatic conditions based on the average across several years and therefore considered much smoother environmental changes than the daily stochastic variation that is observed in outdoor conditions. Similarly, these studies do not take into account biotic interactions such as competition, herbivory and pathogen attacks that may trigger various flowering-time responses[121–123]. The next challenge in identifying genes underlying ecologically relevant traits in *A. thaliana* will certainly be the phenotyping of plants that have established themselves in natural populations without human interference.

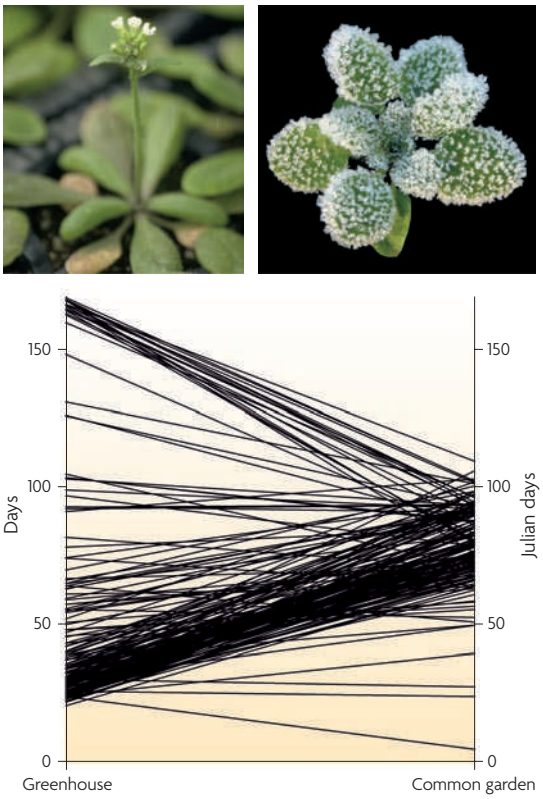


Figure 4 | **Reaction norms of flowering time between the greenhouse and the common garden.** *Arabidopsis thaliana* reveals extensive genotype–environment interactions between greenhouse and outdoor conditions. Flowering time has been scored for 183 worldwide accessions in greenhouse conditions (20 °C, 16 hour photoperiod)[17] and in a common garden at the University of Lille (Northern France)[23]. In the greenhouse, flowering time is expressed in days since sowing. In the common garden, seeds were sown in late September and flowering time is expressed in Julian days since 1 January. Note that most accessions flowered in early spring, that is, late March, in the common garden. Images courtesy of B. Brachi, Université des Sciences et Technologies de Lille.

Genotype–environment interactions. Like many other plant species with a worldwide distribution, *A. thaliana* can be found in contrasting habitats. Phenotypic plasticity might thus be a key factor in the process of adaptation to newly colonized geographical areas[124,125]. The selfing reproductive system of *A. thaliana* enables one to replicate genotypes within and between environments, allowing direct examination of phenotypic plasticity. Such studies have revealed extensive genetic variation for reaction norms, suggesting the occurrence of strong G×E interactions in *A. thaliana*[123,126] (FIG. 4).

Work has begun to dissect the genetic architecture of G×E interactions in *A. thaliana* for various environments, such as seasons, water availability, nitrogen sources and plant density[90,116,122,127–132]. That said, the molecular and mechanistic bases of the functional polymorphisms underlying G×E interactions remain poorly known. A recent and large experiment to phenotype *A. thaliana* mutants that are impaired in particular flowering-time pathways has started to fill this gap by phenotyping plants in common gardens located in different geographical regions[24]. The authors demonstrated that early flowering conferred by loss-of-function alleles at the *FRIGIDA* (*FRI*) gene was negated by a shift of a few days in germination in early autumn. Very recently, 473 *A. thaliana* accessions were phenotyped for flowering time across two planting seasons in each of two simulated local climates (Spain and Sweden) in growth chambers[133]. In this study, all 12 flowering time QTLs detected by GWA mapping showed sensitivity to seasonal planting and/or simulated local climate. Other GWA mapping experiments performed in multiple geographic regions, such as the flowering time studies that constitute the Ecological Genomics of *Arabidopsis*

Phenotypic plasticity
The ability of an organism to develop a phenotypic state, depending on its external and internal environment.

Reaction norm
The set of phenotypes expressed by a genotype under different environmental conditions.

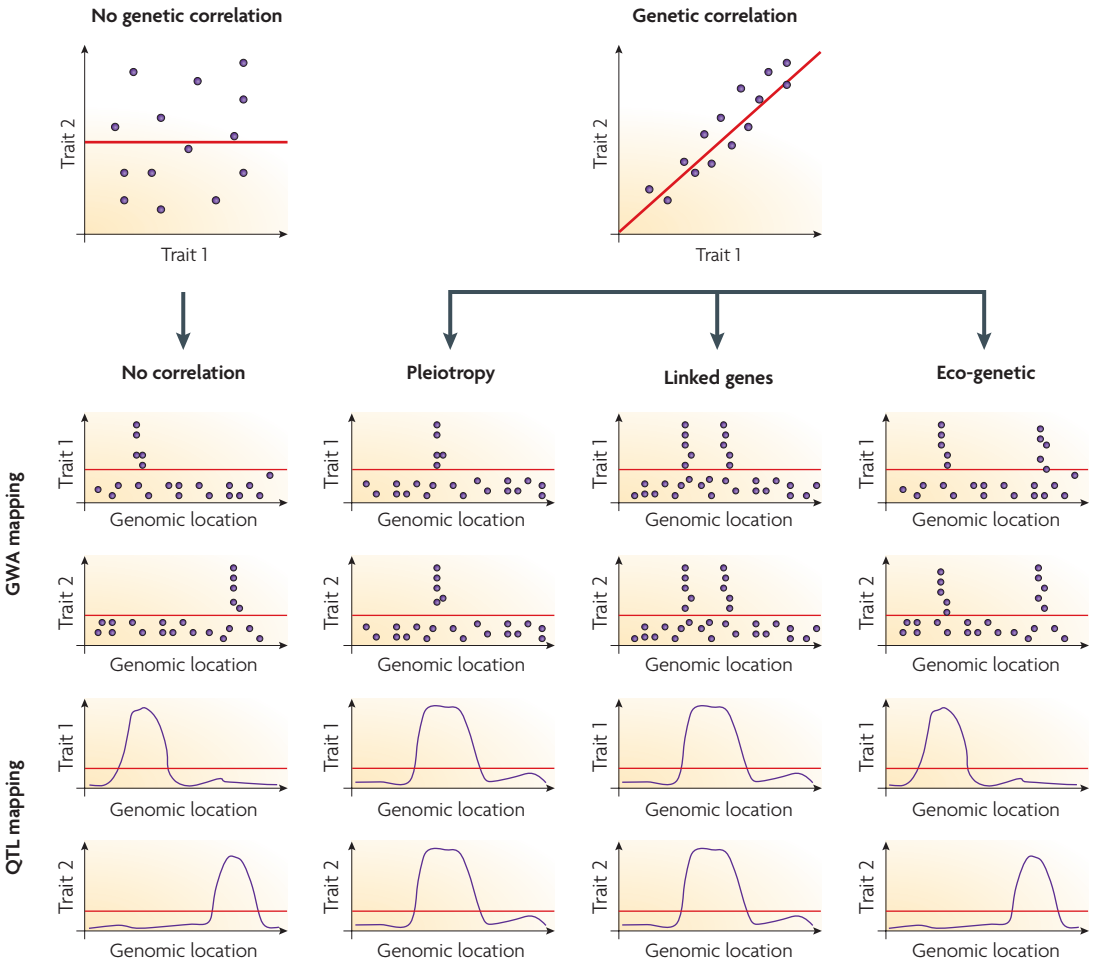


Figure 5 | **Unravelling the origin of genetic correlations.** Both genome-wide association (GWA) and quantitative trait locus (QTL) mapping will reveal distinct genomic locations that are associated with natural variation of two uncorrelated phenotypic traits. Genetic correlations might originate from pleiotropic genes, from physically linked genes or from distinct genes that have been selected by covarying ecological factors. Pleiotropy: the same genomic location is identified by GWA mapping for both traits and is overlapped by the same QTL region. Linked genes: physically close genomic locations are identified by GWA mapping for both traits and are overlapped by only one QTL region. Eco-genetic: physically distant genomic locations are identified by GWA mapping for both traits. Each genomic location is overlapped by only one QTL region.

Development (EGAD) project, are expected to make further significant advances in the understanding of G × E interactions in *A. thaliana*. As extensive year-to-year variation in seed production of the same *A. thaliana* genotype has been detected at one Swedish field station across 8 years[22], GWA mapping experiments will also be usefully replicated across successive years.

Quantitative traits network. Individuals are simultaneously confronted with multiple selective pressures, leading to selection for a global phenotypic optimum that results from trade-offs among specific traits[26,134]. For a selfing annual such as *A. thaliana*, seed production — a proxy of fitness in *A. thaliana* — results from a combination of morphological, physiological, phenological and life history traits (BOX 1). In humans, this corresponds to clinical outcomes that can be thought of as a synthesis of intermediate phenotypes (that is, risk factors)[135]. Path analyses have been used to describe the phenotypic networks that underlie fitness in plants[119,136,137], thereby assessing direct and indirect selection on individual traits. Performing statistical estimation of correlated genome associations[86] may provide insight into the process of adaptation by unravelling the origin of genetic correlations among phenotypic traits, that is, pleiotropy versus genetically linked genes[138].

Still, genetic correlations may also originate from joint selection of covarying ecological factors[139] (FIG. 5). For example, several phenotypic traits in *A. thaliana* are correlated with latitude. Whereas the decrease in solar radiation that is associated with latitude might be thought to select on relative growth rate (RGR)[140], precipitation and/or temperature related to latitude might be the key climatic factors that act on bolting time[100]. In the case of independent genetic bases for correlated traits, crossing two accessions with extreme phenotypes should enable one to break down the genetic correlation observed among accessions (FIG. 5). Thus, whereas

Path analysis
A statistical method that provides estimates of the magnitude and significance of causal relationships between two or more variables.

Pleiotropy
The effect of a gene on more than one phenotypic trait.

GWA mapping will identify the same genomic regions associated with correlated traits, traditional linkage mapping may help to distinguish the origin of genetic correlations.

Conclusion and perspectives

GWA mapping clearly facilitates the identification of genes associated with natural variation in phenotypic traits. In *A. thaliana,* it is also relatively easy to identify false positives and negatives through the strategic combination of traditional linkage and association mapping[23,52], something that is not feasible in humans and many other systems. In addition, the identification of common alleles of major effect suggests a relatively simple genetic architecture for many adaptive traits in *A. thaliana*; such results have not been apparent in maize, mice, flies and humans, in which many loci of small effect have been detected[141]. It remains to be determined whether this is due to the focal species or to focal traits. After functional polymorphisms are validated, it is possible to study the history of selection for these polymorphisms[87,142] and then determine the main contributors to adaptation, that is, new mutations versus standing genetic variation.

Performing ecological genomics by adding ecology to the studies of phenotype–genotype associations will forge a better understanding of adaptation in *A. thaliana*[143], enabling us to retrace the trajectory of adaptive traits in natural populations[2,3] and potentially improve crop yield and quality. Soon, the current revolution in NGS technologies will additionally facilitate ecological genetics in non-model plant species.

Although GWA mapping gives access to the unit of evolution — that is, the gene — the unit of selection — that is, the phenotype — must not be forgotten. Indeed, the next frontier in GWA mapping is high-throughput phenotyping. Due to the development of NGS technologies, genomic resources are rapidly accumulating, but phenotypic data collected in a natural context remain scarce. Automated platforms have been recently developed for phenotyping in growth chambers[144,145], and an International Plant Phenomics Network (IPPN) was recently set up to provide new technologies for high-throughput phenotyping. As genetic variation is exposed to natural selection in nature, such automated platforms are desperately needed to allow phenotyping of plants in natural conditions.

1. Fisher, R. A. (ed.) *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. Hermisson, J. & Pennings, P. S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352 (2005).
3. Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
4. Kopp, M. & Hermisson, J. Adaptation of a quantitative trait to a moving optimum. *Genetics* **176**, 715–719 (2007).
5. Kopp, M. & Hermisson, J. The genetic basis of phenotypic adaptation I: fixation of beneficial mutations in the moving optimum model. *Genetics* **182**, 233–249 (2009).
6. Stern, D. L. & Orgogozo, V. Is genetic evolution predictable? *Science* **323**, 746–751 (2009).
An interesting review on the predictability of genetic evolution, with a special emphasis on the factors that influence the distribution of mutations relevant for phenotypic evolution.
7. Rafalski, J. A. Association genetics on crop improvement. *Curr. Opin. Plant Biol.* **13**, 1–7 (2010).
8. Erickson, D. L., Fenster, C. B., Stenoien, H. K. & Price, D. Quantitative trait locus analyses and the study of evolutionary process. *Mol. Ecol.* **13**, 2505–2522 (2004).
9. Mitchell-Olds, T. & Schmitt, J. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* **441**, 947–952 (2006).
10. Ellegren, H. & Sheldon, B. C. Genetic basis of fitness differences in natural populations. *Nature* **452**, 169–175 (2008).
11. Bergelson, J., Stahl, E., Dudek, S. & Kreitman, M. Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics* **148**, 1311–1323 (1998).
12. Le Corre, V. Variation at two flowering time genes within and among populations of *Arabidopsis thaliana*: comparison with markers and traits. *Mol. Ecol.* **14**, 4181–4192 (2005).
13. Bomblies, K. *et al.* Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet.* **6**, e10000890 (2010).
14. Platt, A. *et al.* The scale of population structure in *Arabidopsis thaliana*. *PLoS Genet.* **6**, 1–8 (2010).
References 13 and 14 describe the scale and patterns of genetic variability in natural populations of *A. thaliana*, using either local stands or worldwide samples, respectively.
15. Koornneef, M., Alonso-Blanco, C. & Vreugdenhil, D. Naturally occurring genetic variation in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* **55**, 141–172 (2004).
16. Alonso, J. M. & Ecker, J. R. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. *Nature Rev. Genet.* **7**, 524–536 (2006).
17. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in a common set of *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
This first report of GWA mapping in plants highlights both advantages and pitfalls related to GWA mapping.
18. Nordborg, M. & Weigel, D. Next-generation genetics in plants. *Nature* **456**, 720–723 (2008).
19. Myles, S. *et al.* Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* **21**, 2194–2202 (2009).
20. Mitchell-Olds, T. Complex-traits analysis in plants. *Genome Biol.* **11**, 113 (2010).
21. Rosenberg, N. A. *et al.* Genome-wide association studies in diverse populations. *Nature Rev. Genet.* **11**, 356–366 (2010).
22. Frenkel, M., Jänkänpää, H. J. & Jansson, S. An illustrated gardener's guide to transgenic *Arabidopsis* field experiments. *New Phytol.* **180**, 545–555 (2008).
23. Brachi, B. *et al.* Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* **6**, e1000940 (2010).
The first report of dual linkage–GWA mapping in a common garden, strengthening evidence for the need to use complementary methods to decrease both false-positive and false-negative rates in *A. thaliana*.
24. Wilczek, A. M. *et al.* Effects of genetic perturbation on seasonal life history plasticity. *Science* **323**, 930–934 (2009).
This outstanding paper links functional genomics and ecologically realistic conditions for a better understanding of selection on flowering-time genes in *A. thaliana*.
25. Thomas, D. Gene–environment-wide association studies: emerging approaches. *Nature Rev. Genet.* **11**, 259–272 (2010).
26. Roff, D. A. Contributions of genomics to life-history theory. *Nature Rev. Genet.* **8**, 116–125 (2007).
27. Lister, R., Gregory, B. D. & Ecker, J. R. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr. Opin. Plant Biol.* **12**, 107–118 (2009).
28. Metzker, M. L. Sequencing technologies - the next generation. *Nature Rev. Genet.* **11**, 31–46 (2010).
A well-illustrated review of NGS technologies.
29. Delseny, M., Han, B. & Hsing, Y. I. High throughput DNA sequencing: the new sequencing revolution. *Plant Sci.* **179**, 407–422 (2010).
30. Kowalski, S. P., Lan, T. H., Feldmann, K. A. & Paterson, A. H. QTL mapping of naturally-occurring variation in flowering time of *Arabidopsis thaliana*. *Mol. Genet. Genomics* **245**, 548–555 (1994).
31. Kover, P. X. *et al.* A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* **5**, e1000551 (2009).
32. Lynch, M. & Walsh, S. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1998).
33. Price, A. H. Believe it or not, QTLs are accurate! *Trends Plant Sci.* **11**, 213–216 (2006).
34. Borevitz, J. & Chory, J. Genomics tools for QTL analysis and gene discovery. *Curr. Opin. Plant Biol.* **7**, 132–136 (2004).
35. Tuinstra, M. R., Ejeta, G. & Goldsbrough, P. B. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *Theor. Appl. Genet.* **95**, 1005–1011 (1997).
36. Keurentjes, J. J. B. *et al.* Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. *Genetics* **175**, 891–905 (2007).
37. Roosens, N. H., Willems, G. & Saumitou-Laprade, P. Using *Arabidopsis* to explore zinc tolerance and hyperaccumulation. *Trends Plant Sci.* **13**, 208–215 (2008).
38. Verbruggen, N., Hermans, C. & Schat, H. Molecular mechanisms of metal hyperaccumulation in plants. *New Phytol.* **181**, 759–776 (2009).
39. Schneeberger, K. *et al.* SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* **6**, 550–551 (2009).
40. Laitinen, R. A., Schneeberger, K., Jelly, N. S., Ossowski, S. & Weigel, D. Identification of a spontaneous frame shift mutation in a nonreference *Arabidopsis* accession using whole genome sequencing. *Plant Physiol.* **153**, 652–654 (2010).
41. Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039–1042 (2010).
42. Ehrenreich, I. M. *et al.* Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics* **183**, 325–335 (2009).
43. Zhu, C., Gore, M., Buckler, E. S. & Yu, J. Status and prospects of association mapping in plants. *Plant Genome* **1**, 5–20 (2008).

44. Kim, S. *et al.* Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* **39**, 1151–1155 (2007).
45. Zhang, X., Richards, E. J. & Borevitz, J. O. Genetic and epigenetics dissection of *cis* regulatory variation. *Curr. Opin. Plant Biol.* **10**, 142–148 (2007).
46. Aranzana, M. J. *et al.* Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* **1**, e60 (2005).
47. Warren, R. F., Henk, A., Mowery, P., Holub, E. & Innes, R. W. A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene *RPS5* partially suppresses multiple bacterial and downy mildew resistance genes. *Plant Cell* **10**, 1439–1452 (1998).
48. Grant, M. R. *et al.* Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
49. Nemri, A. *et al.* Genome-wide survey of *Arabidopsis* natural variation in downy mildew resistance using combined association and linkage mapping. *Proc. Natl Acad. Sci. USA* **107**, 10302–10307 (2010).
50. Todesco, M. *et al.* Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* **465**, 632–636 (2010).
51. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
52. Zhao, K. *et al.* An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet.* **3**, e4 (2007).
53. Manenti, G. *et al.* Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. *PLoS Genet.* **5**, e1000331 (2009).
54. Dillmann, C., Bar-Hen, A., Guérin, D., Charcosset, A. & Murigneux, A. Comparison of RFLP and morphological distances between maize *Zea mays* L. inbred lines. Consequences for germplasm protection purposes. *Theor. Appl. Genet.* **95**, 92–102 (1997).
55. Vignieri, S. N., Larson, J. G. & Hoekstra, H. E. The selective advantage of crypsis in mice. *Evolution* **64**, 2153–2158 (2010).
56. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. A single amino-acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101–104 (2003).
A well-designed study to functionally validate the genetic basis of an adaptive trait in a non-model species.
57. Veyrieras, J.-B., Goffinet, B. & Charcosset, A. MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinformatics* **8**, 49–64 (2007).
58. Simon, M. *et al.* Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* **178**, 2253–2264 (2008).
59. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
60. Le Corre, V., Roux, F. & Reboud, X. DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. *Mol. Biol. Evol.* **19**, 1261–1271 (2002).
61. Yan, L. *et al.* The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. *Science* **303**, 1640–1644 (2004).
62. Buckler, E. S. *et al.* The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
An ambitious mapping study using the NAM populations of maize in a set of field experiments that reveals that, unlike in *A. thaliana*, many alleles of small effect mediate flowering time in an additive fashion.
63. Yu, J., Holland, J. B., McMullen, M. D. & Buckler, E. S. Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539–551 (2008).
64. Stich, B. Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* **183**, 1525–1534 (2009).
65. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* **3**, e196 (2005).
66. Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285 (2001).
67. Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671 (1999).
68. Bakker, E., Traw, B. M., Toomajian, C., Kreitman, M. & Bergelson, J. Low levels of polymorphism in genes that control the activation of defense response in *Arabidopsis thaliana*. *Genetics* **178**, 2031–2043 (2008).
69. Weigel, D. & Mott, R. The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol.* **10**, 107 (2009).
70. Caicedo, A. L., Richards, C., Ehrenreich, I. M. & Purugganan, M. Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana MAF2–5* flowering time gene cluster. *Mol. Biol. Evol.* **26**, 699–711 (2009).
71. Kroymann, J., Donnerhacke, S., Schnabelrauch, D. & Mitchell-Olds, T. Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus. *Proc. Natl Acad. Sci. USA* **100**, 14587–14592 (2003).
72. Richards, E. J. Inheritance epigenetic variation — revisiting soft inheritance. *Nature Rev. Genet.* **7**, 395–401 (2006).
73. Bossdorf, O., Richards, C. L. & Pigliucci, M. Epigenetics for ecologists. *Ecol. Lett.* **11**, 106–115 (2008).
74. Vaughn, M. W. *et al.* Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5**, e174 (2007).
75. Johannes, F. *et al.* Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* **5**, e10000530 (2009).
References 74 and 75 demonstrate the importance of epigenetic alterations in *A. thaliana* as a possible source of heritable phenotypic variation and the need to epigenotype natural accessions to infer causal relationships between genotype and phenotype.
76. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 1–14 (2008).
77. Zhang, X., Shiu, S., Cal, A. & Borevitz, J. O. Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tilling arrays. *PLoS Genet.* **4**, e1000032 (2008).
78. Laird, P. W. Principles and challenges of genome-wide DNA methylation analysis. *Nature Rev. Genet.* **11**, 191–203 (2010).
79. Sillanpää, M. J. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity* 14 Jul 2010 (doi: 10.1038/hdy.2010.91).
80. Kang, H. M. *et al.* Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723 (2008).
81. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nature Rev. Genet.* **11**, 459–463 (2010).
82. El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A. J. M., Raz, V. & Koornneef, M. A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. *Nature Genet.* **29**, 435–440 (2001).
83. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature Genet.* **42**, 355–360 (2010).
84. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **29**, 2633–2635 (2007).
85. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348–354 (2010).
86. Kim, S. & Xing, E. P. Statistical estimation of correlated genome associations to a quantitative trait network. *PLoS Genet.* **5**, e1000587 (2009).
87. Tishkoff, S. A. *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genet.* **39**, 31–40 (2007).
88. Puniyani, K., Kim, S. & Xing, E. P. Multi-population GWA mapping via multi-task regularized regression. *Bioinformatics* **26**, i208–i216 (2010).
This paper describes the development of a promising multi-population GWA mapping method that enables the detection of causal genetic markers that are unique to a subset of the populations.
89. O'Malley, R. C. & Ecker, J. R. Linking genotype to phenotype using the *Arabidopsis* unimutant collection. *Plant J.* **61**, 928–940 (2010).
90. Weinig, C. *et al.* Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. *Genetics* **162**, 1875–1884 (2002).
The first paper describing QTL mapping in outdoor conditions. It makes clear that phenotypes should be assessed in ecologically realistic conditions to allow the detection of genes underlying natural variation in *A. thaliana*.
91. Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77 (2003).
92. Vergunst, A. C. & Hooykaas, P. J. Cre/lox-mediated site-specific integration of Agrobcaterium T-DBA in *Arabidopsis thaliana* by transient expression of cre. *Plant Mol. Biol.* **38**, 393–406 (1998).
93. Alonso-Blanco, C. *et al.* What has natural variation taught us about plant development, physiology, and adaptation? *Plant Cell* **21**, 1877–1896 (2009).
94. Egli, D. B. Seed-fill duration and yield of grain crops. *Adv. Agron.* **83**, 243–279 (2004).
95. Kover, P. X. & Schaal, B. A. Genetic variation for disease resistance and tolerance among *Arabidopsis thaliana* accessions. *Proc. Natl Acad. Sci. USA* **99**, 11270–11274 (2002).
96. Gao, L., Roux, F. & Bergelson, J. Quantitative fitness effects of infection in a gene-for-gene system. *New Phytol.* **184**, 485–494 (2009).
97. Levins, R. *Evolution in Changing Environments* (Princeton Univ. Press, New Jersey, 1968).
98. Becker, U., Dostal, P., Jorritsma-Wienk, L. D. & Matthies, D. The spatial scale of adaptive population differentiation in a wide-spread, well-dispersed plant species. *Oikos* **117**, 1865–1976 (2008).
99. Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M. & Purugganan, M. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl Acad. Sci. USA* **101**, 15670–15675 (2004).
100. Stinchcombe, J. R. *et al.* A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. *Proc. Natl Acad. Sci. USA* **101**, 4712–4717 (2004).
101. Marquis, R. in *Plant Resistance to Herbivores and Pathogens: Ecology, Evolution and Genetics* (eds Fritz, R. S. & Simms, E. L.) 301–325 (Univ. Chicago Press, Illinois, 1992).
102. Stratton, D. A. & Bennington, C. C. Measuring spatial variation in natural selection using randomly-sown seeds of *Arabidopsis thaliana*. *J. Evol. Biol.* **9**, 215–228 (1996).
103. Goss, E. M. & Bergelson, J. Fitness consequences of pathogen infection of *Arabidopsis thaliana* with its natural bacterial pathogen *Pseudomonas viridiflava*. *Oecologia* **152**, 71–81 (2007).
104. Mani, G. S. Evolution of resistance in the presence of two insecticides. *Genetics* **109**, 761–783 (1985).
105. Roux, F., Paris, M. & Reboud, X. Delaying weed adaptation to herbicide by environmental heterogeneity: a simulation approach. *Pest Manag. Sci.* **64**, 16–29 (2008).
106. Kassen, R. & Bell, G. Experimental evolution in *Chlamydomonas*. IV. Selection in environments that vary through time at different scales. *Heredity* **80**, 732–741 (1998).
107. Kassen, R. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**, 173–190 (2002).
108. Bell, G. Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Philos. Trans. R. Soc. Lond. B* **365**, 87–97 (2010).
109. Lennartsson, T., Tuomi, J. & Nilsson, P. Evidence for an evolutionary history of overcompensation in the grassland biennial *Gentianella campestris* (*Gentianaceae*). *Am. Nat.* **149**, 1147–1155 (1997).
110. Poveda, K., Steffan-Dewenter, I., Scheu, S. & Tscharntke, T. Effects of below- and above-ground herbivores on plant growth, flower visitation and seed set. *Oecologia* **135**, 601–605 (2003).
111. Lefebvre, V., Kiani, S. P. & Durand-Tardif, M. A focus on natural variation for abiotic constraints response in the model species *Arabidopsis thaliana*. *Int. J. Mol. Sci.* **10**, 3547–3582 (2009).
112. Wielgolaski, F. E. Phenological modifications in plants by various edaphic factors. *Int. J. Biometeorol.*. **45**, 196–202 (2001).
113. Nord., E. A. & Lynch, J. P. Delayed reproduction in *Arabidopsis thaliana* improves fitness in soil with suboptimal phosphorus availability. *Plant Cell Environ.* **31**, 1432–1441 (2008).

114. Baxter, I. *et al.* A coastal cline in sodium accumulation in *Arabidopsis thaliana* is driven by natural variation of the sodium transporter AtHKT1-1. *PLoS Genet.* (in the press).
115. Gardner, K. M. & Latta, R. G. Identifying loci under selection across contrasting environments in *Avena barbata* using quantitative trait locus mapping. *Mol. Ecol.* **15**, 1321–1333 (2006).
116. Weinig, C. *et al.* Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics* **165**, 321–329 (2003).
117. Malmberg, R. L., Held, S., Waits, A. & Mauricio, R. Epistasis for fitness-related quantitative traits in *Arabidopsis thaliana* grown in the field and in the greenhouse. *Genetics* **171**, 2013–2027 (2005).
118. Li, Y., Roycewicz, P., Smith, E. & Borevitz, J. O. Genetics of local adaptation in the laboratory: flowering time quantitative trait loci under geographic and seasonal conditions in *Arabidopsis*. *PLoS ONE* **1**, e105 (2006).
119. Scarcelli, N., Cheverud, J. M., Schaal, B. A. & Kover, P. X. Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. *Proc. Natl Acad. Sci. USA* **104**, 16986–16991 (2007).
120. Kover, P. X. *et al.* Pleiotropic effects of environment-specific adaptation in *Arabidopsis thaliana*. *New Phytol.* **183**, 816–825 (2009).
121. Dorn, L. A., Pyle, E. H. & Schmitt, J. Plasticity to light cues and resources in *Arabidopsis thaliana*: testing for adaptive value and costs. *Evolution* **54**, 1982–1994 (2000).
122. Weinig, C., Stinchcombe, J. R. & Schmitt, J. QTL architecture of resistance and tolerance traits in *Arabidopsis thaliana* in natural environments. *Mol. Ecol.* **12**, 1153–1163 (2003).
123. Roux, F., Gao, L. & Bergelson, J. Impact of initial pathogen density on resistance and tolerance in a polymorphic disease resistance gene system in *Arabidopsis thaliana*. *Genetics* **185**, 283–291 (2010).
124. Kingsolver, J. G., Pfennig, D. W. & Servedio, M. R. Migration, local adaptation and the evolution of plasticity. *Trends Ecol. Evol.* **17**, 540–541 (2002).
125. Weinig, C. & Schmitt, J. Environmental effects on the expression of quantitative trait loci and implications for phenotypic evolution. *Bioscience* **54**, 627–635 (2004).
126. Donohue, K. *et al.* Environmental and genetic influences on the germination of *Arabidopsis thaliana* in the field. *Evolution* **59**, 740–757 (2005).
127. Kliebenstein, D., Figuth, A. & Mitchell-Olds, T. Genetic architecture of plastic methyl jasmonate responses in *Arabidopsis thaliana*. *Genetics* **161**, 1685–1696 (2002).
128. Rauh, B. L., Basten, C. & Buckler, E. S. Quantitative trait loci analysis of growth response to varying nitrogen sources in *Arabidopsis thaliana*. *Theor. Appl. Genet.* **104**, 743–750 (2002).
129. Loudet, O., Chaillou, S., Krapp, A. & Daniel-Vedele, F. Quantitative trait loci analysis of water and anion contents in interaction with nitrogen availability in *Arabidopsis thaliana*. *Genetics* **163**, 711–722 (2003).
130. Ungerer, M. C., Halldorsdottir, S. S., Purugganan, M. D. & Mackay, T. F. Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* **165**, 353–365 (2003).
131. Hausmann, N. J. *et al.* Quantitative trait loci affecting $\delta^{13}C$ and response to differential water availability in *Arabidopsis thaliana*. *Evolution* **59**, 81–96 (2005).
132. Botto, J. F. & Coluccio, M. P. Seasonal and plant-density dependency for quantitative trait loci affecting flowering time in multiple populations of *Arabidopsis thaliana*. *Plant Cell Environ.* **30**, 1465–1479 (2007).
133. Li, Y., Huang, Y., Bergelson, J., Nordborg, M. & Borevitz, J. Association mapping of local climate sensitive QTL in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* (in the press).
134. Mackay, T. F., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature Rev. Genet.* **10**, 565–577 (2009). **A comprehensive Review of the consensus and challenges for obtaining a better understanding of the genetic architecture of complex phenotypic traits.**
135. Carlson, C. S., Eberle, M. A., Kruglyak, L. & Nickerson, D. A. Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446–452 (2004).
136. Bergelson, J. The effects of genotype and the environment on costs of resistance in lettuce. *Am. Nat.* **143**, 349–359 (1994).
137. Byers, D. L. Evolution in heterogeneous environments and the potential of maintenance of genetic variation in traits of adaptive significance. *Genetica* **123**, 107–124 (2005).
138. Gardner, K. M. & Latta, R. G. Shared quantitative trait loci underlying the genetic correlation between continuous traits. *Mol. Ecol.* **16**, 4195–4209 (2007).
139. Armbruster, W. S. & Schwaegerle, K. E. Causes of covariation of phenotypic traits among populations. *J. Evol. Biol.* **6**, 261–276 (1996).
140. Li, B., Suzuki, J.-I. & Hara, T. Latitudinal variation in plant size and relative growth rate in *Arabidopsis thaliana*. *Oecologia* **115**, 293–301 (1998).
141. Flint, J. & Mackay, T. F. C. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
142. Toomajian, C. *et al.* A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS Biol.* **4**, e137 (2006).
143. Ungerer, M., Johnson, L. C. & Herman, M. A. Ecological genomics: understanding gene and genome function in the natural environment. *Heredity* **100**, 178–183 (2008).
144. Jansen, M. *et al.* Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct. Plant Biol.* **11**, 902–914 (2009).
145. Massonnet, C. *et al.* Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiol.* **152**, 2142–2157 (2010).
146. Hagenblad, J. & Nordborg, M. Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. *Genetics* **161**, 289–298 (2002).
147. Ehrenreich, I. M., Stafford, P. A. & Purugganan, M. The genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations *vs.* quantitative trait locus mapping. *Genetics* **173**, 1223–1236 (2007).
148. McMullen, M. D. *et al.* Genetic properties of the maize nested association mapping population. *Science* **325**, 737–740 (2009).
149. Kusterer, B. *et al.* Analysis of triple testcross design with recombinant inbred lines reveals a significant role for epistasis in heterosis for biomass-related traits in *Arabidopsis*. *Genetics* **175**, 2009–2017 (2007).
150. Kusterer, B. *et al.* Heterosis for biomass-related traits in *Arabidopsis* investigated by quantitative trait loci analysis of the triple testcross design with recombinant inbred lines. *Genetics* **177**, 1839–1850 (2007).
151. Shindo, C., Lister, C., Crevillen, P., Nordborg, M. & Dean, C. Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. *Genes Dev.* **20**, 3079–3083 (2006).
152. Darvasi, A. & Soller, M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207 (1995).
153. Balasubramanian, S. *et al.* QTL mapping in new *Arabidopsis thaliana* advanced intercross-recombinant inbred lines. *PLoS ONE* **4**, e4318 (2009).
154. Loudet, O., Gaudon, V., Trubuil, A. & Daniel-Vedele, F. Quantitative trait loci controlling root growth and architecture in *Arabidopsis thaliana* confirmed by heterogeneous inbred family. *Theor. Appl. Genet.* **110**, 742–753 (2005).

Acknowledgements

The authors give special thanks to M. Horton and B. Brachi for stimulating discussions on placing GWA mapping studies in an ecological context, to O. Loudet for links to automated platforms of phenotyping and to E. Xing for links to the GenAMap platform for structured GWA mapping. We are grateful for funding from the US National Science Foundation (MCB-0603515), the US National Institutes of Health (GM083068) and the French l'Agence Nationale de la Recherche (NT09_473214).

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Joy Bergelson's homepage: http://bergelson.uchicago.edu
1001 Genomes Project: http://1001genomes.org
Arabidopsis Tiling Array information (Borevitz laboratory): http://borevitzlab.uchicago.edu/resources/computational-resources/arabidopsis-tiling-array-info
Arabidopsis Tiling Array information (Nordborg laboratory): http://walnut.usc.edu/2010/data/250k-data-version-3.04
Ecological Genomics of *Arabidopsis* Development: http://www.egad.ksu.edu
GenAMap (an integrated analytic and visualization platform for eQTL and GWA study analysis): http://cogito-b.ml.cmu.edu/genamap
Genomic analysis of the genotype–phenotype map: http://www.gmi.oeaw.ac.at/en/research/magnus-nordborg/genomic-analysis-of-the-genotypephenotype-map/
International Plant Phenomics Network: http://www.plantphenomics.com/index.php?index=2
***Nature Reviews Genetics* series on Genome-wide association studies:** http://www.nature.com/nrg/series/gwas/index.html
***Nature Reviews Genetics* series on Study designs:** http://www.nature.com/nrg/series/studydesigns/index.html
RegMap lines: http://bergelson.uchicago.edu/a.thaliana-resources
Results of GWA studies for 107 traits: http://cypress.usc.edu/DisplayResults

ALL LINKS ARE ACTIVE IN THE ONLINE PDF